

# REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

AFRL-SR-AR-TR-04-

0098

|   |   |  |   |
|---|---|--|---|
| 1. AGENCY USE ONLY (Leave blank)  |   | 2. REPORT DATE<br>January 30, 2004                         | 3. REPORT TYPE AND I<br>Final Technical Report      |
| 4. TITLE AND SUBTITLE<br><br>Auditory Peripheral Processing of Degraded Speech  |   |  | 5. FUNDING NUMBERS<br><br>F49620-03-1-0173          |
| 6. AUTHOR(S)<br><br>Oded Ghitza   |   |  |   |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br><br>Sensimetrics Corporation<br>48 Grove Street<br>Somerville, MA 02144   |   |  | 8. PERFORMING ORGANIZATION<br>REPORT NUMBER         |
| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br><br>NK   |   |  | 10. SPONSORING / MONITORING<br>AGENCY REPORT NUMBER |
| 11. SUPPLEMENTARY NOTES   |   |  |   |
| 12a. DISTRIBUTION / AVAILABILITY STATEMENT<br><br>Unrestricted distribution<br><br>DISTRIBUTION STATEMENT A<br>Approved for Public Release<br>Distribution Unlimited  |   |  |   |
| 13. ABSTRACT (Maximum 200 Words)<br><br>This is a final report for a stand-alone grant supporting the first 9 months of a 4-year research program entitled "Auditory peripheral processing of degraded speech". The underlying thesis is that the auditory periphery contributes to the robust performance of humans in speech reception in noise through a concerted contribution of the efferent feedback system (which stabilizes the operating point of the cochlea) and post-auditory-nerve functions (such as acoustic-edge detectors). The 4-year program includes psychophysical experiments aimed at documenting the role of these two peripheral components; developing a computational model; and using the emerging model as a front-end to an automatic speech recognition system. The proposed timetable allocates the first 12 months for psychophysical experiments. This report describes efforts to assess the role of the MOC efferent system in speech reception in the presence of sustained background noise, by using a test of initial consonant reception (the Diagnostic Rhyme Test) performed by subjects with normal hearing. Activation of selected parts of the efferent system was attempted by presenting speech and noise in various configurations (gated/continuous, monaural/binaural). Initial results support the hypothesis of a significant efferent contribution to initial phone discrimination in noise. |   |  |   |
| 14. SUBJECT TERMS   |   |  | 15. NUMBER OF PAGES<br>13                           |
|   |   |  | 16. PRICE CODE                                      |
| 17. SECURITY CLASSIFICATION<br>OF REPORT<br>Unclassified  | 18. SECURITY CLASSIFICATION<br>OF THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION<br>OF ABSTRACT<br>Unclassified | 20. LIMITATION OF ABSTRACT                          |

20040220 313

**DISTRIBUTION STATEMENT A**

Approved for Public Release

Distribution Unlimited

**A. INTRODUCTION**

Presently we don't know which auditory processing principles are behind the level of performance humans demonstrate in the task of speech reception in noise. One way to improve our capabilities in designing speech-processing systems that will be effective in the presence of background noise is to advance our understanding of how the auditory periphery operates in such environments, and to translate this understanding into a computational model. Our definition of the auditory periphery is restricted to the processing which takes place prior to lexical access, on speech segments that are as long as 100ms (i.e., as long as the duration of a dyad<sup>1</sup>). Currently, we have a reasonable understanding of the processing principles in the ascending auditory pathway up through the auditory nerve (AN) [e.g., the cochlea; the inner hair cells (IHC)] and an increasing understanding of brain-stem nuclei (such as the cochlear nucleus, the superior olivary complex, the inferior colliculus). We have limited understanding of the descending pathway [mainly the MOC to outer-hair-cells (OHC) feedback] and very little understanding of how the ascending and the descending pathways interact.

A key observation is that human performance in tasks related to speech intelligibility deteriorates only modestly with worsening environmental conditions even for tasks with a minimal cognitive load (i.e. no contextual information is available, and the only source of information available to the listener are the AN firing patterns). In contrast, simulated AN representations – generated by state-of-the-art auditory models – deteriorate at a much faster rate. One may attribute the robust performance by the human to the existence of (1) mechanisms that stabilize the AN firing patterns in the presence of noise, resulting in a representation of acoustic-phonemic information which is more consistent with the representation in a quiet background, and (2) efficient peripheral post-AN mechanisms that are capable of extracting important acoustic-phonemic cues even from noisy AN patterns. Our underlying assumption is that the stabilizing mechanism and the post-AN mechanisms work in concert in providing the observed graceful degradation of human performance in noise. We suggest that the success of post-AN mechanisms in reliably extracting speech-related information in noise is partly due to the "stabilizing" effect of the efferent system. Current models of the periphery are based upon the ascending pathway up through the AN. We envision a model of the periphery that utilizes the role of the descending pathway, and the way the ascending and the descending pathways interact.

One auditory mechanism that may play a role in regulating cochlear mechanics is the medial olivocochlear (MOC) efferent feedback system. This report describes our efforts at quantifying the possible role of this system in speech reception in the presence of background noise.

**B. BACKGROUND****B.2 MOC efferents: morphology and physiology**

Numerous papers have been published providing detailed morphological and neurophysiological description of the MOC efferent feedback system (e.g., Gifford and Guinan, 1983; Guinan, 1996; Kawase and Liberman, 1993; Liberman, 1988; Liberman and Brown 1986; May and Sachs, 1992; Warr, 1978; Winslow and Sachs, 1988). MOC efferents originate from neurons medial, ventral and anterior to the medial superior olivary nucleus (MSO), have myelinated axons, and terminate directly on Outer Hair Cells (OHC). Medial efferents project predominantly to the contralateral cochlea, the innervation is largest near the center of the cochlea, with the crossed innervation biased toward the base compared to the uncrossed innervation (e.g., Guinan, 1996). Roughly two-third of medial efferents respond to ipsilateral sound, one-third to contralateral sound, and a small fraction to sound in either ear. Medial efferents have tuning curves that are similar to, or slightly wider than, those of AN fibers (e.g., Liberman and Brown 1986), and they project to different places along the cochlear partition in a tonotopical manner. Finally, medial efferents have longer latencies and group delays than AN fibers. In response to tone or noise bursts, most MOC efferents have latencies of 10-40ms. Group delays measured from modulation transfer functions are much more tightly clustered, averaged at about 8ms (Gummer *et al.*,

<sup>1</sup> An acoustic segment from the midpoint of one phoneme to the midpoint of the adjacent phoneme

1988). We currently do not have a clear understanding of the functional role of this mechanism. Few suggestions have been offered, such as shifting of sound-level functions to higher sound levels, antimasking effect on responses to transient sounds in a continuous masker, preventing damage due to intense sound (e.g., Guinan, 1996). One speculated role, which is of particular interest for this proposal, is a dynamic regulation of the cochlear operating point depending on background acoustic stimulation, resulting in robust human performance in perceiving speech in a noisy background (e.g., Kiang *et al.*, 1987). There are a few neurophysiological studies to support this role. Using anesthetized cats with noisy acoustic stimuli, Winslow and Sachs (1988), for example, showed that by stimulating the MOC nerve bundle electrically, the dynamic range of discharge rate at the AN is partly recovered. Measuring neural responses of awake cats to noisy acoustic stimuli, May and Sachs (1992) showed that the dynamic range of discharge rate at the AN level is only moderately affected by changes in levels of background noise.

### **B.3 MOC efferents: psychophysics – speech and speech-like stimuli**

Few behavioral studies indicate the potential role of the MOC efferent system in perceiving speech in the presence of background noise. Dewson (1968) presented evidence that MOC lesions impair the abilities of monkeys to discriminate the vowel sounds [i] and [u] in the presence of masking noise but have no effect on the performance of this task in quiet. More recently, Giraud *et al.* (1997), and Zeng *et al.* (2000) showed that the performance of humans with severed MOC feedback deteriorates phoneme perception when the speech is presented in a noisy background. Hienz *et al.* (1998) confirmed the classical findings of Dewson (1968) and the more recent observations of Giraud *et al.* (1997); they showed that de-efferentation of the cochlea produces vowel discrimination deficits in cats, particularly when performance was measured in the presence of high levels of background noise.

## **C. PRESENT STUDY**

We collected data by psychophysical experiments using subjects with normal hearing. A controlled activation of selected parts of the efferent system was achieved by the use of judiciously designed stimulus conditions.

### **Phone discrimination in noise – subjects with normal hearing**

We conducted phoneme discrimination experiments using speech with a minimal context, hence focusing on the role of the auditory periphery (by reducing the role of higher auditory layers to a minimum). Toward this end, we used the Diagnostic Rhyme Test (DRT; Voiers, 1983), which uses the one-interval two-alternative forced-choice paradigm as the administrative procedure. The properties of the DRT are reviewed in Sec. C.1.a. Our assumptions about the functioning of the MOC efferent system in the presence of sustained background noise, which guided the design considerations of the proposed experiments, are summarized in Sec. C.1.b. The definition of the resulting stimulus conditions and the rationale behind their design are presented in Sec. C.1.c. The results and data analysis are summarized in Sec. C.1.d.

#### **C.1.a Database and experimental procedure – Voiers DRT**

The DRT (Diagnostic Rhyme Test) was suggested by Voiers (1983) as a way of measuring the intelligibility of processed speech. From an acoustic point of view, Voiers' DRT database covers initial dyads of spoken CVCs. The database consists of 96 pairs of confusable words spoken in isolation. Words in a pair differ only in their initial consonants. The dyads are equally distributed among 6 acoustic-phonetic distinctive features and among 8 vowels (hence 2 word-pairs per a [feature×vowel] bin). The feature classification (outlined in Table 1) follows the binary system suggested by Jakobson, Fant and Halle (Jakobson *et al.*, 1952), and the vowels are [ee] and [it] (High-Front), [eh] and [at] (High-Back), [oo] and [oh] (Low-Front) and [aw] and [ah] (Low-Back). In our version of the DRT the vowels are collapsed into 4 quadrants (High-Front, High-Back, Low-Front, Low-Back), hence 4 word-pairs per a [feature×quadrant] bin. Driven by considerations stemming from efferent activation properties (see Sec. C.1.c, topic 1.1), we truncated the duration of the consonantal part of each word stimuli to a maximum of 50ms (measured from the time-instant of consonantal-to-vocalic transition, backwards) as illustrated in Fig. 1. Consequently, the sound quality of some word stimuli was degraded. The word-pairs associated

with these words were removed from the database, resulting in a database of 72 word-pairs evenly distributed (i.e. 3 word-pairs per a [feature $\times$ quadrant] bin).

The psychophysical procedure is very carefully controlled, to assure a task with minimum cognitive load. The listeners are well trained and are very familiar with the database, including the voice quality of the individual speakers. The experiment is a one-interval two-alternative forced-choice experiment. First, the subject is presented visually with a pair of rhymed words. Then, one word of the pair (selected at random) is presented aurally and the subject is required to indicate which of the two words was played. This procedure is repeated until all the words in the database have been presented. In our version of the DRT words were played sequentially, one every 2.5 seconds; the visual presentation preceded the aural presentation by 1sec., and the decision (binary) was made within 1sec. of the aural presentation. Words in the database were divided into sessions, and the overall duration of one session was limited to about 2.5 minutes.

### C.1.b Stimulus conditions and presumed efferent activity – background

The design of the stimuli for the proposed experiments was guided by the observed behavior of the MOC efferent system in the presence of sustained background noise, as illustrated in Fig. 2. The ipsilateral ear is defined as the ear that will be presented with the speech signal. For monaural presentation of noise to the ipsilateral ear (left panel), the information pathway relevant to MOC efferent activation consists of the auditory nerve projection to the posteroventral subdivision of the cochlear nucleus (PVCN); the MOC reflex interneurons at the PVCN, likely to be chopper units (Brown *et al.*, 2003), which project, across midline, to MOC neurons at the medial superior olive (MSO) whose axons cross back (via the COCB nerve bundle) to the ipsilateral cochlea, innervating the outer hair cells (OHCs). Important neurophysiological observations are: (1) MOC neurons project to OHC in a tonotopic manner, (2) The effect on the OHC is about 100ms after stimulus onset, and (3) the activity of the MOC neurons in background noise is sustained (i.e. non-significant rate adaptation, Brown *et al.*, 2003). For monaural presentation of noise to the contralateral ear (right panel), the information pathway consists of the reflex interneurons at the PVCN at the contralateral cochlear nucleus, which project, across midline, to MOC neurons at the medial superior olive (MSO) whose axons project to the ipsilateral cochlea. An important observation is that the strength of the feedback, in terms of number of MOC neurons excited by monaural stimuli, is roughly 2:1 in favor of the ipsilateral pathway (e.g. Liberman, 1988). It is assumed that for binaural presentation of noise the feedback strength sums to 1<sup>2</sup>.

### C.1.c Stimulus conditions and presumed efferent activity – definition

The stimulus conditions are illustrated in Fig. 3 and the presumed efferent activity they invoke are summarized in Table 2.

#### 1. Efferent activation.

- 1.1. *Monaural presentation.* Fig. 3(a) illustrates the baseline condition, [Sm-Gated], where the noise turns on and off in sync with the word stimulus. The noise duration is 1 sec. long, the gap between two successive words is 2.5 seconds (all word stimuli in the database are less than 1 sec. in duration). In the DRT, words in a given word-pair differ in the initial CV dyad. We limited the duration of the consonantal part of the dyad to a maximum of 50ms (see Sec. C.1.a). Recalling that the effect on the MOC efferent on the OHC is about 100ms after stimulus onset, the 50ms long consonantal part and the immediate 50ms long vocalic, coarticulated part of the initial CV dyad are, therefore, presumed to be processed by a cochlea with no efferent elicitation. This is marked by "0" ipsilateral activation in Table 2, row 1. In contrast, in condition [Sm-Cont.] (Fig. 3(b)) noise remains on throughout the session, hence eliciting ipsilateral activation (with strength of  $\frac{2}{3}$ , as indicated in Table 2, row 2).

<sup>2</sup> A noteworthy observation is the existence MOC neurons with binaural inputs (Liberman, 1988); the behavior of these units is less understood.



- 1.2. *Noise characteristics, noise intensity and SNR.* We used Speech-Shape noise, additive. Calculation of SNR was over the first 100ms of the vowel (starting at the consonantal-to-vocalic transition point of the initial dyad). Noise intensity, in dB SPL, was the parameter and was fixed throughout a session. The other parameter in the experiment was the SNR; the intensity of the word stimuli was adjusted – amplified, or attenuated, by a constant – to satisfy the nominal SNR value. The range of noise intensities and SNRs will be detailed in Sec. C.1.d.
  - 1.3. *Binaural presentation.* The signal presented to the ipsilateral ear is the same as in the monaural presentation. The contralateral ear receives noise alone, with the same intensity as the noise in the ipsilateral ear. This is illustrated in Figs. 3(c) and 3(d) (for [Sm-Gated] and [Sm-Cont.] in the ipsilateral ear, respectively). Presumed efferent elicitation is indicated in rows 3 and 4 of Table 2.
2. Effect of binaural processing. A question arises whether the expected contribution to performance should be attributed to the efferent system *per se*, or to an advantage resulting from processing by binaural mechanisms up stream. To address this question, in all conditions that involve a presentation to both ears (Table 2, rows 3 through 7) we used two kinds of contralateral noise: [Nu] – noise in both ears is uncorrelated (no binaural advantage); and [No] – same noise realization in both ears (possible binaural advantage).
  3. Effect of integration in ascending pathway. As we found out in Experiment I [Sec. C.1.d, Table 3(a)], performance for the binaural conditions deteriorated compared to the corresponding monaural conditions. This result seems unexpected, since an additional contralateral efferent elicitation is supposed to improve performance. One possible hypothesis to explain this oddity is a possible degradation in representing speech cues in the ascending pathway as a result of stimulating the contralateral ear with noise only. Adding noise to the contra ear, in addition to the noisy speech presented to the ipsilateral ear, may create a noisier integrated image of the speech input. To resolve this issue we introduced condition [Sm-Cont.]–[Nu-Gated] (Fig. 3(f)), where the noise in the contra ear is turned off during the time-interval where speech is presented in the ipsi ear, resulting in speech image provided by the ipsilateral path alone. If our hypothesis about the role of the efferent system is correct, the performance of this condition (i.e. condition [Sm-Cont.]–[Nu-Gated]) should be at least as good as condition [Sm-Cont.] (i.e. monaural ipsi activation).
  4. [Sm-Gated] vs. [Sm-Cont.] – fused vs. segregated auditory images? As we found out in Experiment I (Sec. C.1.d), performance for condition [Sm-Cont.] (ipsilateral elicitation) is significantly better than that measured for the baseline condition [Sm-Gated]. A question arises whether this advantage is due to the efferent system *per se*, or the effect of involvement of auditory segregation mechanisms; a word stimulus added to a continuously running noise (condition [Sm-Cont.]) may be segregated easier than having gated-noise and a word stimulus switch on at the same time instant. To resolve this issue we introduced the monaural condition [Sm-Gated-WGN] (Fig. 3(e)), where the silence gaps of condition [Sm-Gated] are filled with white Gaussian noise, with the same intensity as the baseline noise. We assume that, due to the markedly different characteristics of the two noise realizations, the image of the noisy speech is as fused as in condition [Sm-Gated]. However, the efferent elicitation is as in condition [Sm-Cont.]. (This is indicated in row 8 of Table 2.) If our hypothesis about the role of the efferent system is correct, the performance of this condition (i.e. condition [Sm-Gated-WGN]) should be as good as condition [Sm-Cont.].

### C.1.d Results and data analysis

The study comprises two experiments, with Experiment II emerging out of Experiment I.

In Experiment I we measured the performance for four conditions, no efferent activation (the baseline condition [Sm-Gated], illustrated in Fig. 3(a)), monaural, ipsilateral efferent activation (condition [Sm-Cont.], Fig. 3(b)), and two conditions of binaural efferent activation [conditions [Sm-Cont.]–[Nu] and [Sm-Cont.]–[No], Fig. 3(d)], one with uncorrelated noise in the contralateral ear, one with the same noise in both ears (i.e., providing binaural advantage). Four subjects were tested (on their left ear). Their performance is presented in Table 3(a) per subject, since performance is expected to vary across

subjects due to differences in the strength of their efferent reflex (e.g. Guinan *et al.*, 2003). A table-entry shows the mean word-error<sup>3</sup>  $\pm$  its standard deviation. The table-entry also contains, in parentheses, the t-statistic for the difference between means, relative to the baseline condition. For the number of degrees-of-freedom in the data here, a difference between the means is significant at the  $p < 0.05$  level, if the absolute value of the t-statistic is (roughly) greater than 2.5.

The results show a significant reduction in the mean number of errors for the ipsilateral efferent activation compared to no efferent activation, for all subjects (at least 25%). Using the t-test criteria, the difference between the means (ipsilateral condition vs. baseline) is significant. To our surprise, adding continuous contralateral noise resulted in degradation in performance compared to the ipsilateral (monaural) efferent activation condition, consistent across subjects (rows 3 and 4). This was the case for either uncorrelated noise to the contra ear, or same noise to both ears (the later should provide a binaural advantage). From efferent elicitation viewpoint, these results seem unexpected since adding noise to the contra ear should strengthen the efferent response to the ipsilateral cochlea and should, therefore, improve performance<sup>4</sup>. (The t-test here indicates insignificant difference between the means, compared to baseline). Two more observations are noteworthy. First, as was expected performance varies across subjects, but the trend as a function of efferent elicitation condition is similar. Second, question arises whether the advantage in the ipsilateral condition is due to the efferent system *per se*; it may be that words presented to the ipsilateral ear in the presence of continuous noise are segregated easier than having gated noise and a word stimulus switch on at the same time instant (as in the baseline condition).

In Experiment II we addressed two questions that emerged out of Experiment I, (1) what is the reason for the degradation in performance for binaural efferent activation, and (2) is the advantage for ipsilateral efferent activation due to differences in the nature of the projected auditory images (i.e. "fused" vs. segregated)? For the first question we hypothesize that the reason for the drop in performance is a further corruption of speech cues in the ascending pathway, resulted from presenting extra noise (with no extra signal) to the contra ear. To test this hypothesis we turned the noise in the contra ear off during the time-interval when speech was presented to the ipsi ear [condition [Sm-Cont.]-{Nu-Gated}, Fig. 3(f)]. Results suggest restoration of the performance level, closer to the level of performance measured for the ipsi efferent activation<sup>5</sup>. The second question was addressed by creating a monaural signal where the added noise during the duration of the word stimuli is speech-shaped, while the noise in between the word stimuli is white [condition [Sm-Gated-WGN], Fig. 3(e)]. In this way the ipsi efferent elicitation is maintained while the image of the noisy word stimulus is fused (due to the different nature of the noise segments). Results suggest restoration of the performance level, closer to the level of performance measured for the ipsi efferent activation.

Finally, errors collected in DRT sessions can be averaged over subjects and plotted as a function of the Jakobsonian acoustic-phonetic dimensions, as illustrated in Figures 4(a) and 4(b). The panels at each figure represent different efferent elicitation conditions. Fig. 4(a) is for word-pairs in the High-Front vowel quadrant, and Fig. 4(b) is for word-pairs in the Low-Front vowel quadrant. Using knowledge about the acoustic correlates of the Jakobsonian dimensions, the information in Fig. 4 can be used to identify the origins of the errors in the time-frequency plane. Notice the difference in error distributions between Fig. 4(a) and Fig. 4(b), and the difference between panels within each figure. These differences may result from different efferent effects depending on the location of the formants (different in each vowel quadrant).

<sup>3</sup> The range of the mean word-error is [0,3].

<sup>4</sup> The degradation in performance when adding noise to the contralateral ear is also unexpected from the binaural processing viewpoint. Results from detection studies indicate little or no additional masking from a contralateral noise. When the contralateral noise is in-phase, in fact, there is an improvement in detection (Durlach and Colburn, 1978). The absence of improvement here may be due to the fact that the cues for consonant discrimination tend to be above 1500 Hz, where binaural detection improvements are negligible.

<sup>5</sup> Only two subjects were tested in Experiment II.

## BIBLIOGRAPHY

- Brown *et al.* (2003). Responses of medial olivocochlear (MOC) neurons: Specifying the central pathways of the MOC reflex. Submitted for publication.
- Dewson, J. H. (1968). Efferent olivocochlear bundle: some relationships to stimulus discrimination in noise. *J. Neurophysiol.*, 31:122–130.
- Durlach, N.I. and Colburn, H.S. (1978). "Binaural phenomena," in *Handbook of Perception, Vol. IV.* edited by E.C. Carterette and M.P. Friedman (Academic Press, New York).
- Gifford, M. L. and Guinan, J. J. (1983). Effects of crossed olivocochlear bundle stimulation on cat auditory-nerve responses to tones. *J. Acoust. Soc. Am.*, 74:115–123.
- Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., and Chery-Croze, S. (1997). Auditory efferents involved in speech-in-noise intelligibility. *NeuroReport*, 8:1799-1783.
- Godfrey, D., Kiang, N., and Norris B. (1975). Single unit activity in the posteroventral cochlear nucleus. *J. Comput. Neurosci.* 162:247–268.
- Guinan, J. J. (1996). Physiology of Olivocochlear Efferents. In Dallos, P., Popper, A. N. and Fay, R. R., editors, *The Cochlea*, pages 435–502, Springer, New-York.
- Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). Medial olivocochlear efferent reflex in humans: Otoacoustic Emission (OAE) measurement issues and the advantages of Stimulus Frequency OAEs. *J.A.R.O.* 4, In Press.
- Gummer, M., Yates, G. K., and Johnstone, B. M. (1988). Modulation transfer function of efferent neurons in the guinea pig cochlea. *Hear. Res.*, 36:41–52.
- Hienz, R.D., Stiles, P., and May, B. J. (1998). Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats. *Hear. Res.*, 116:10–20.
- Jakobson, R., Fant, C. G. M., and Halle, M. (1952). Preliminaries to speech analysis: the distinctive features and their correlates. Technical report, Acoustic Laboratory, Massachusetts Institute of Technology.
- Kawase, T. and Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones. *J. Neurophysiol.*, 70:2519:2532.
- Kiang, N. Y. S., Guinan, J. J., Liberman, M. C., Brown, M. C., and Eddington, D. K. (1987). Feedback control mechanisms of the auditory periphery: implication for cochlear implants. In Banfai, P., editor, *International Cochlear Implant Symposium*. Duren, West Germany.
- Liberman, M. C. (1988). Responses properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise. *J. Neurophysiol.*, 60:1779–1798.
- Liberman, M. C. and Brown, M. C. (1986). Physiolog and anatomy of single olivocochlear neurons in the cat. *Hear. Res.*, 24:17–36.

May, B. J. and Sachs, M. B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. *J. Neurophysiol.*, 68:1589–1603.

Voiers, W. D. (1983). Evaluating processed speech using the diagnostic rhyme test. *Speech Technology*, 1(4):30–39.

Warr, W. B. (1978). The olivocochlear bundle: its origins and terminations in the cat. In Naunton, R. and Fernandez, C., editors, *Evoked Electrical Activity in the Auditory Nervous System*, pages 43–63. Academic Press, New York.

Wiederhold, M. and Kiang, N. Y. S. (1970). Effects of electrical stimulation on crossed olivocochlear bundle on single auditory-nerve fibers in the cat. *J. Acoust. Soc. Am.*, 48:950–965.

Winslow, R. L. and Sachs, M. B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear. Res.*, 35:165–190.

Zeng, P. G., Martino, K. M., Linthcum, F. H., and Soli, S. (2000). Auditory perception in vestibular neurectomy subjects. *Hear. Res.*, 142:102–112.



Table 1. Samples of word-pairs used in Voiers' DRT (1983)

| Voicing (VC)<br>( <i>Voiced - Unvoiced</i> )          | Nasality (NS)<br>( <i>Nasal - Oral</i> )   | Sustention (ST)<br>( <i>Sustained - Interrupted</i> ) |
|---|--|---|
| veal - feel<br>zed - said<br>-                        | meat - beat<br>neck - deck<br>-            | vee - bee<br>fence - pence<br>-                       |
| Sibilation (SB)<br>( <i>Sibilated - Assibilated</i> ) | Graveness (GV)<br>( <i>Grave - Acute</i> ) | Compactness (CM)<br>( <i>Compact - Diffuse</i> )      |
| cheep - keep<br>jot - got<br>-                        | peak - teak<br>wad - rod<br>-              | key - tea<br>got - dot<br>-                           |

Table 2. Presumed efferent activation. See Fig. 3 for description of signal condition.

|   | Fig . 4 | Stimulus condition    | Ipsi activation | Contra activation |
|---|---------|-----------------------|-----------------|-------------------|
| 1 | (a)     | [Sm-Gated]            | 0               | 0                 |
| 2 | (b)     | [Sm-Cont.]            | $\frac{2}{3}$   | 0                 |
| 3 | (c)     | [Sm-Gated]-[Nu]       | 0               | $\frac{1}{3}$     |
| 4 | (d)     | [Sm-Cont.]-[Nu]       | $\frac{2}{3}$   | $\frac{1}{3}$     |
| 5 | (d)     | [Sm-Cont.]-[No]       | $\frac{2}{3}$   | $\frac{1}{3}$     |
| 6 | (f)     | [Sm-Cont.]-[Nu-Gated] | $\frac{2}{3}$   | $\frac{1}{3}$     |
| 7 | (f)     | [Sm-Cont.]-[No-Gated] | $\frac{2}{3}$   | $\frac{1}{3}$     |
| 8 | (e)     | [Sm-Gated-WGN]        | $\frac{2}{3}$   | 0                 |

Table 3(a). Experiment I. Mean word errors (out of 3 words)  $\pm$  Standard deviation. In parentheses are the t-value of difference between the means relative to baseline condition (row 1). Columns - subjects.

| Presumed efferent activation               | EB (d.f.=16)               | JZ (d.f.=16)              | YG (d.f.=10)              | OG (d.f.=10)              |
|--|----------------------------|---------------------------|---------------------------|---------------------------|
| No activation                              | 0.36 $\pm$ 0.08            | 0.38 $\pm$ 0.09           | 0.41 $\pm$ 0.05           | 0.53 $\pm$ 0.11           |
| Ipsilateral activation                     | 0.25 $\pm$ 0.05<br>(3.52)  | 0.28 $\pm$ 0.04<br>(3.32) | 0.29 $\pm$ 0.06<br>(3.74) | 0.31 $\pm$ 0.04<br>(4.91) |
| Binaural activation,<br>uncorrelated noise | 0.42 $\pm$ 0.04<br>(-2.02) | 0.37 $\pm$ 0.06<br>(0.52) | 0.36 $\pm$ 0.04<br>(2.06) | 0.37 $\pm$ 0.06<br>(3.35) |
| Binaural, same noise<br>in both ears       | 0.34 $\pm$ 0.06<br>(0.49)  | 0.32 $\pm$ 0.05<br>(1.89) | 0.39 $\pm$ 0.08<br>(0.61) | 0.35 $\pm$ 0.08<br>(3.44) |

Table 3(b). Experiment II.

| Presumed efferent activation             | - | JZ (d.f.=16)              | - | OG (d.f.=10)              |
|--|---|---------------------------|---|---------------------------|
| Binaural, gated noise in<br>contra ear   | - | 0.33 $\pm$ 0.03<br>(1.61) | - | 0.31 $\pm$ 0.04<br>(4.75) |
| Ipsilateral activation,<br>"fused" image | - | 0.32 $\pm$ 0.03<br>(1.16) | - | 0.29 $\pm$ 0.08<br>(4.63) |

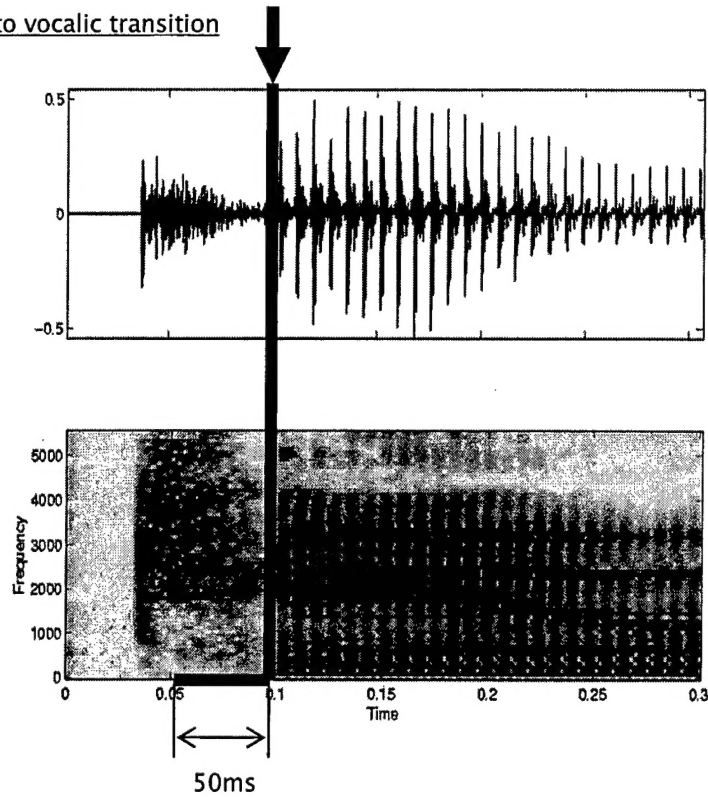
Consonantal to vocalic transition

Figure 1. The consonantal part of each word stimuli was truncated to a maximum of 50ms (measured from the time-instant of consonantal to vocalic transition; backwards).

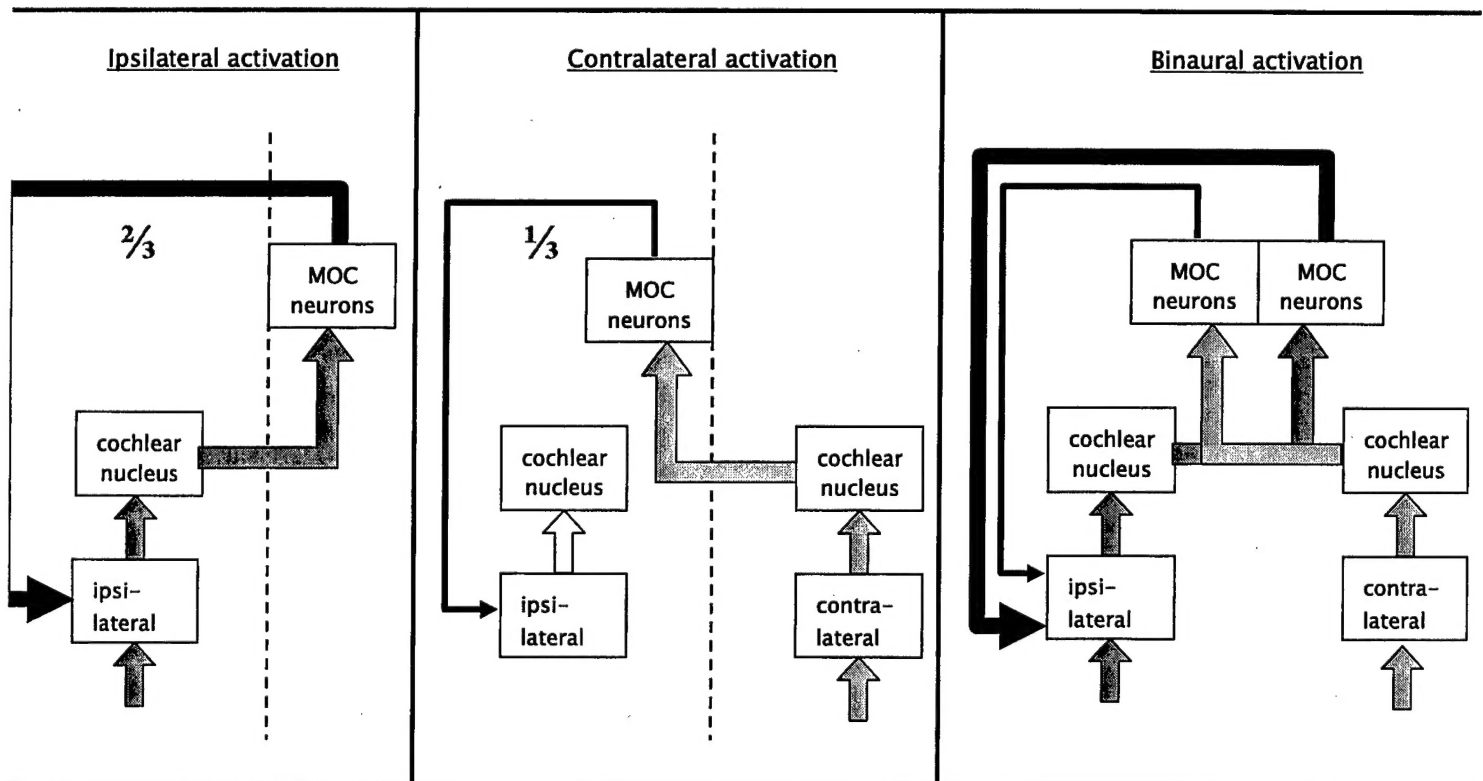


Figure 2. Illustration of the MOC neurons wiring – a block diagram

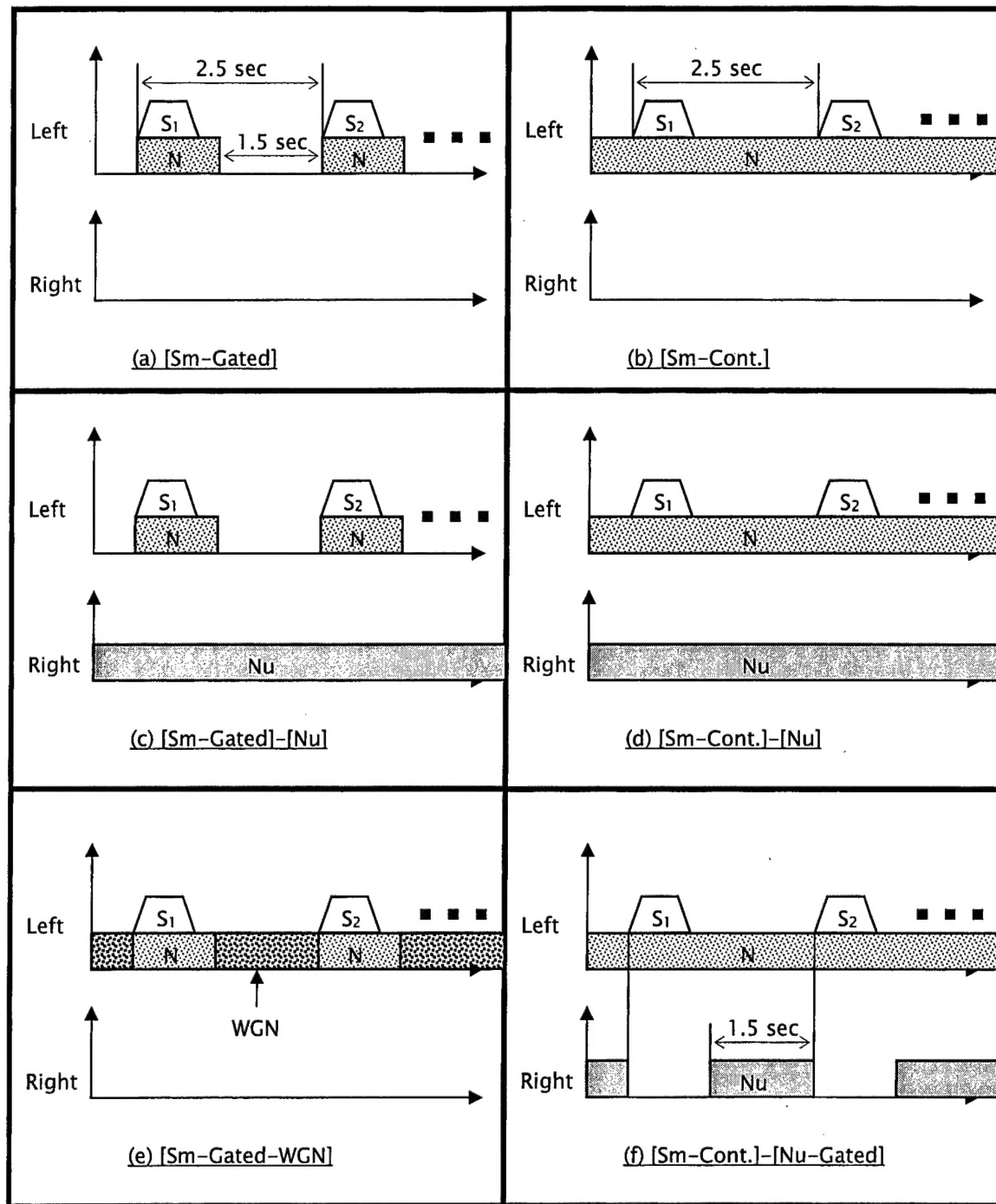


Figure 3. Illustration of signal conditions. Sm stands for Signal-monaural; N stands for noise at ipsi ear; Nu stands for noise at contralateral ear, uncorrelated to noise at ipsilateral ear; No - correlated noise at cotralateral ear (N, Nu and No are speech-shaped noise signals); WGN stands for white Gaussian noise.



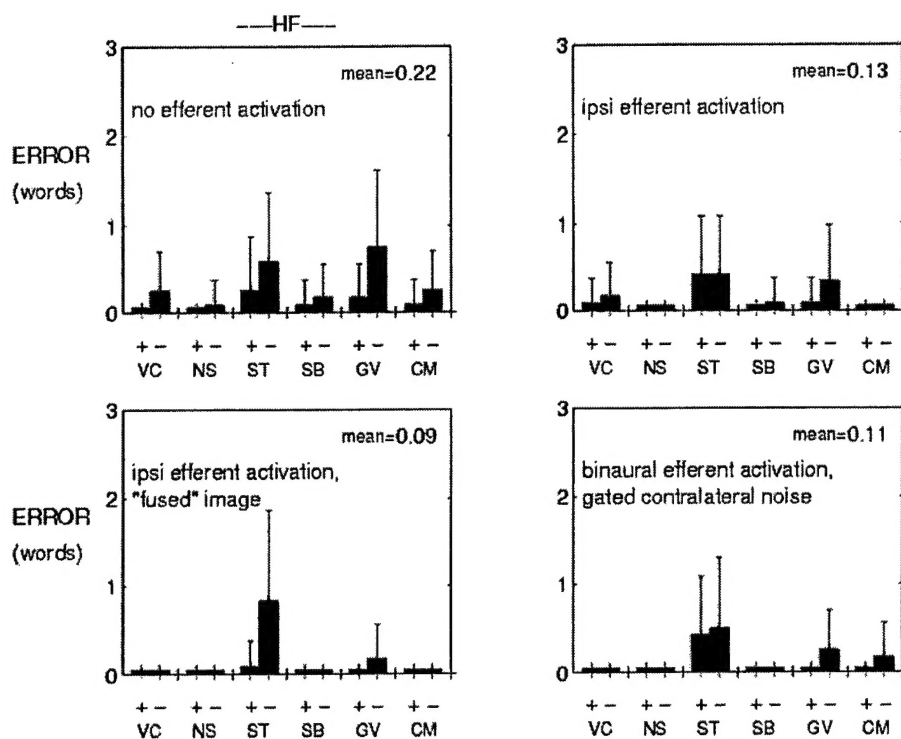


Figure 4(a). Error patterns for High-Front vowels in the DRT database. Noise is speech-shaped, with intensity of 60dB SPL. SNR is 5dB. Abscissa - Jakobsonian acoustic features (+ for attribute present, - for attribute absent). Ordinate - mean errors, in words (out of 3 words per bin).

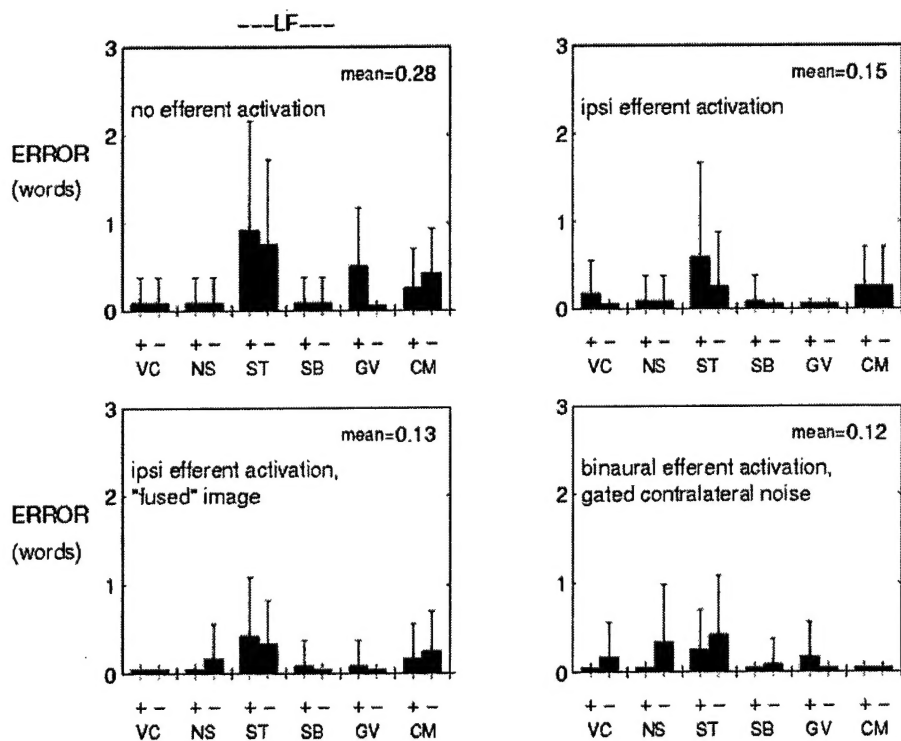


Figure 4(b). Error patterns for Low-Front vowels in the DRT database.